Title: Derivation and external validation of risk algorithms for cerebrovascular (re)hospitalisation in patients with type 2 diabetes: two cohorts study

Authors: Dahai Yu[1,2], Yamei Cai[1], Jonathan Graffy[3], Daniel Holman[4], Zhanzheng Zhao[1], David Simmons[5]

1Department of Nephrology, the First Affiliated Hospital, Zhengzhou University, Zhengzhou 450052, China.

2Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Keele ST5 5BG, UK.

3Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, Cambridgeshire CB2 0SR, United Kingdom

4Department of Sociological Studies, University of Sheffield, Sheffield S10 2TU, United Kingdom

5Western Sydney University, Campbelltown, Sydney NSW 2760, Australia

Correspondence: Zhanzheng Zhao
Department of Nephrology
the First Affiliated Hospital, Zhengzhou University
Zhengzhou 450052, CHINA
Tel +86 139 3852 5666
Fax +86 371 6698 8753
Email zhanzhengzhao@zzu.edu.cn

Correspondence: David Simmons
Macarthur Clinical School
School of Medicine, Western Sydney University
Locked Bag 1797, Campbelltown NSW 2760, AUSTRALIA
Tel (61+2) 4620 3899
Fax (61+2) 4620 3890
Email dsworkster@gmail.com

Word count: 2,953 words

Short title: Risk score for CeV (re)hospitalisation in T2DM

ABSTRACT

Aims: Cerebrovascular disease is one of more typical reasons for hospitalisation and re-hospitalisation in people with type 2 diabetes. We aimed to derive and externally validate two risk prediction algorithms for cerebrovascular hospitalisation and re-hospitalisation.

Methods: Two independent cohorts were used to derive and externally validate the two risk scores. The development cohort comprises 4,704 patients with type 2 diabetes registered in 18 general practices across Cambridgeshire. The validation cohort includes 1,121 type 2 patients from a post-trial cohort data. Outcomes were cerebrovascular hospitalisation within two years and cerebrovascular re-hospitalisation within ninety days of the previous cerebrovascular hospitalisation. Logistic regression was applied to derive the two risk scores for cerebrovascular hospitalisation and re-hospitalisation from development cohort, which were externally validated in the validation cohort.

Results: The incidence of cerebrovascular hospitalisation and re-hospitalisation was 3.76% and 1.46% in the development cohort, and 4.99% and 1.87% in the external validation cohort. Age, gender, body mass index, blood pressures, and lipid profiles were included in the final model. Model discrimination was similar in both cohorts, with all C-statistics > 0.70, and very good calibration of observed and predicted individual risks.

Conclusion: Two new risk scores that quantify individual risks of cerebrovascular hospitalisation and re-hospitalisation have been well derived and externally validated. Both scores are on the basis of a few of clinical measurements that are commonly available for patients with type 2 diabetes in primary care settings and could work as tools to identify individuals at high risk of cerebrovascular hospitalisation and re-hospitalisation.

Keywords: Cerebrovascular disease; Diabetes population; Risk prediction; Primary care

INTRODUCTION

Type 2 Diabetes as a risk factor for cerebrovascular diseases has been found to be markedly associated with increased risk of cerebrovascular mortality. One meta-analysis revealed that in comparison with people without diabetes, people with diabetes had a 2.27-fold of increased risk of cerebrovascular disease [1]. As cerebrovascular disease is one of the major causes of death and disability in people with type 2 diabetes [2], risk algorithms to predict cerebrovascular disease have been increasingly developed to facilitate the effective management of high risk individuals [3].

It is common for people with diabetes to be admitted to hospital, with one in five inpatients having diabetes in some age groups in England [4]. Cerebrovascular diseases is one of the more common causes for hospitalisation in patients with type 2 diabetes [5]. And it is also common for patients with type 2 diabetes to be re-hospitalised for cardiovascular or cerebrovascular disease [6]. The associated increased inpatient costs are marked factors to the health burden borne by heath care system as a result of diabetes and often reflects manageable morbidities suffered by patients with diabetes. A prediction tool to identify individuals at particularly high risk of cerebrovascular hospitalisation and re-hospitalisation would facilitate subsequent more intensive interventions.

A systematic review identified 12 risk scores to predict coronary heart or cerebrovascular disease conducted in patients with type 2 diabetes [7]. However among the 12 risk scores only two were developed for stroke and neither had external validation [7]. So far, there have been no prediction models developed for cerebrovascular disease in people with type 2 diabetes. Furthermore there have been no models derived and validated to predict cerebrovascular hospitalisation and re-hospitalisation in type 2 diabetes patients.

The objective of this study was to derive and externally validate new risk prediction algorithms based on reliable ordinary clinical measurements recorded in primary care settings for cerebrovascular hospitalisation within the following two years and cerebrovascular re-hospitalisation within 90 days of a prior cerebrovascular hospitalisation.

MATERIAL AND METHODS

Data setting and study population

Two prospective cohorts derived from Cambridgeshire, the United Kingdom were utilised in this study. The derivation cohort included primary care electronic health record data and was used to derive risk algorithms to predict cerebrovascular hospitalisation and re-hospitalisation. The external validation cohort included post-trial data and was utilised to externally validate the two risk algorithms.

Derivation cohort

The derivation cohort included type 2 diabetes patients registered in 18 general practices across Cambridgeshire, England, in 2008/2009 with linkage to inpatient hospitalisation (Secondary Uses Service (SUS)) data as part of a review of diabetes care across Cambridgeshire by the local health board, National Health Service (NHS) Cambridgeshire. Egton Medical Information Systems (EMIS) general practitioner (GP) software system was used in the cohort practices, from which a predefined dataset could be extracted. No systematic selection process for these surgeries was applied, and data extracted were for the whole diabetes population. The follow-up hospitalisation data to 2010–2011 was available to all patients in the derivation cohort. Inpatient hospitalisation to private and NHS hospitals within or outside Cambridgeshire were followed up. Personal identifiers were not released to researchers, and only anonymized datasets were used to conduct all subsequent analyses.

Validation cohort

The validation cohort is a post-trial cohort derived from the RAndomized controlled trial of Peer Support in type 2 Diabetes (RAPSID) [8]. The design and research methods of the RAPSID have been previously published [8]. In brief, RAPSID was designed as a 2x2 factorial cluster randomized controlled trial comparing 4 arms: 1:1 peer support, group peer support, combined support (1:1 plus group peer support) and control in patients with type 2 diabetes. All eligible patients had their type 2 diabetes diagnosed for at least twelve months and those having psychotic illness or dementia were ruled out. Patients were recruited from local communities cross Cambridgeshire and its neighbouring areas of Hertfordshire and Essex. Post-trial follow-up data were only available for patients residence in Cambridgeshire and its neighbouring areas of Hertfordshire that are served by the Cambridgeshire and Peterborough Clinical Commissioning Group (CCG). The intervention was implemented following a pilot in a framework defined by Peers for Progress [9]. The intervention duration was 8-12 months and was concluded between 2 June 2011 to 12 April 2012 [10, 11].

Demographic data, HbA1c, lipid profiles and blood pressure data were collected at baseline. Every eligible patient was followed up till 30 June 2015 (0.91-4.07 years of follow-up from beginning/entry date). Inpatient hospitalisation (NHS & private hospitals), Accident & Emergency (A&E) and outpatient episodes within or outside Cambridgeshire and the included areas of Hertfordshire were collected through Cambridgeshire and Peterborough Clinical CCG [12, 13] and stored as the International Classification of Diseases (ICD-10) codes [14].

Definition of cerebrovascular hospitalisation and re-hospitalisation

The main outcomes in our study are cerebrovascular hospitalisation and rehospitalisation. The cerebrovascular hospitalisation was defined as having ≥1 hospitalisation with cerebrovascular disease (CeVD) as the primary diagnosis (ICD-10: I60–I69 in the first ICD field) over the two-year follow-up and cerebrovascular re-hospitalisation was defined as having ≥1 CeVD re-hospitalisation within ninety days of prior CeVD hospitalisation.

Potential predictors, missing data, and power estimation

Objective clinical measurements including systolic and diastolic blood pressure, body mass index, glycated haemoglobin (HbA1c) and serum lipid profiles were used as predictors in the models to facilitate the external application of the scores. Demographic characters, (sex and age) and whether the patient was prescribed lipid-lowering medicine were also incorporated in our models. In the UK primary care settings, diabetes patients were informed to have their blood pressure and metabolic measurements examined at least once a year since the date of diabetes diagnosis and the most recent measurement was recorded before 1 April 2009 (giving a minimum of fifty days before the first inpatient hospitalisation). The length of diabetes was not commonly recorded, and therefore was not usefully accessible for the model derivation. The specific treatment for diabetes and anti-hypertensive therapy were not accessible in this study. Lipid-lowering prescription was recorded.

Missing information in the derivation cohort included body mass index (3.17%), systolic blood pressure (9.95%), diastolic blood pressure (9.95%), total cholesterol (12.35%), high-density lipoprotein cholesterol (14.56%), and low-density lipoprotein cholesterol (16.27%). Multiple imputation was used to replace missing values by applying a chained equation based on outcome and all potential predictors. 16 imputed datasets were generated for variables with missing values and were then combined over all imputed datasets by Rubin's rule to generate final prediction model estimations.

Few information was missing (<1%) in the external validation cohort and the complete dataset was used in the model validation. Based on 244 cerebrovascular inpatient hospitalisations and 95 cerebrovascular re-hospitalisations and 15 predictors or parameters in the development cohort, an effective sample size (statistical power more than 80% [15]) of 16 cerebrovascular and 6 cerebrovascular re-hospitalisations per predictor or parameters was acquired.

Ethical approval

Ethics approval was granted by the Cambridgeshire REC2 Committee (10/H0308/72), and patients signed-off consent included their agreement for access to inpatient hospitalisation information.

Model development and external validation

The incident cerebrovascular hospitalisation after the first ninety days of the incident occurrence of cerebrovascular re-hospitalisation were treated as binary outcome. For each of the 15 candidate predictors or parameters, the Logistic regression was used to estimate the unadjusted odds ratios. For model development, all candidate predictors were initially included in a multivariable adjusted Logistic regression model. Fractional polynomials were utilised to model non-linear relationships between continuous variables and outcomes.

Lowering lipid treatment was excluded from the multivariable Logistic regression model due to its statistical insignificance ($P>0.1$ for log likelihood) through backward elimination. The eliminated predictor was reinserted into the final prediction models to further examine whether it changed to be statistically significant. Fractional polynomial parameters were also rechecked and re-estimated them if necessary. The risk algorithms were then formed for predicting the log odds of cerebrovascular hospitalisation and cerebrovascular re-hospitalisation by using the Logistic model regression coefficients multiplied by the parameters included in the models together with the intercept terms. This process generated equations for the predicted individual risk=1/(1+e-riskscore), whether the "risk score" is the log odds of cerebrovascular hospitalisation or cerebrovascular re-hospitalisation from the development models.

To facilitate risk score application in primary care, the equations were transferred into risk score charts. The coefficients from the logistic regression were multiplied by 50 and rounded to the nearest integer to generate the score per predictor. Multiplication by 50 was used as the majority of the coefficients was close to an integer, thereby minimizing the rounding effects. The total of prognostic scores indicates the patient probability of cerebrovascular hospitalisation or cerebrovascular re-hospitalisation.

The model performance in terms of the C-statistics and calibration slope (agreement between observed and predicted risks, where 1.00 as ideal) was assessed. The C-statistics indicates the possibility that for any randomly sampled pair of diabetic patients with and without outcomes, the patient with outcomes should have a higher predicted risk [16]. 0.50 of C-statistics indicates no discrimination and 1.00 of calibration slope means perfect discrimination. Optimism (over-fitting) in model performance was corrected through internal validation by bootstrapping 100 samples of the development data. The model development process was then repeated in every bootstrap data to generate a model, applied the model coefficient to the same bootstrap data to quantify apparent performance, and applied the model to the development dataset to examine model performance (C-statistics and calibration slope) and optimism (difference between the apparent and test performance). The overall optimism over all models was then estimated.

Our risk prediction models were applied to individual diabetic patient in the external validation cohort dataset on the basis of the presence of one or more predictors. The final model performance in external validation dataset in terms of discrimination by estimating the C-statistics. We also evaluated model calibration by plotting agreement between observed and predicted probability by decile of the predicted probability.

Stata V15.1 was used for all data analyses. We conducted and presented our study in line with the Transparent Reporting of a multivariate prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [17].

RESULTS

Characteristics of study participants

In the derivation dataset, information of 4,704 type 2 diabetes patients with 244 cerebrovascular hospitalisations within two years and 95 re-hospitalisations within ninety days of a prior cerebrovascular hospitalisation were analysed. The validation dataset incorporated information of 1,121 diabetic patients with 56 cerebrovascular hospitalisations and 21 re-hospitalisations. The baseline characteristics and candidate predictors of the cohorts are presented in **Table**-1. Patients in both cohorts had similar distribution of gender, age, blood pressure and total cholesterol. Patients in the development cohort dataset had a higher level of HbA1c, low-density lipoprotein cholesterol, and high-density lipoprotein cholesterol. Compared with the development cohort dataset, patients in the validation cohort dataset were more likely to take lowering-lipid medicine and had more cerebrovascular hospitalisation and re-hospitalisation.

Model development, performance, and validation

In the development dataset, the absolute risks of cerebrovascular hospitalisation within two years and re-hospitalisation within 90 days post cerebrovascular hospitalisation were 3.76% and 1.46%, respectively. Associations between cerebrovascular hospitalisation and cerebrovascular re-hospitalisation from univariable Logistic regression model are presented in **Supplementary Table-1.**

Among the 10 potential predictors (15 parameters), 9 predictors (12 parameters) were significantly associated with cerebrovascular hospitalisation and re-hospitalisation in our final risk prediction model (**Table-2**). **Table-3** presents apparent and internal validation model performance measurements of the risk prediction model. After the adjustment of optimism, our final risk prediction model was able to discriminate diabetic patients with and without cerebrovascular hospitalisation with a C-statistics of 0.7509 (95% confidence interval 0.7436 to 0.7582), and discriminate diabetic patients with and without cerebrovascular re-hospitalisation with a C-statistics 0.7391 (0.7161 to 0.7451). The agreement between the observed and predicted probability of outcomes showed good apparent calibrations (Top left of **Figure-1** for cerebrovascular hospitalisation and top right of **Figure-1** for cerebrovascular re-hospitalisation). The calibration slope with optimism adjustment was 0.9961 (0.9928 to 0.9995) and 0.9904 (0.9091 to 1.0525) for cerebrovascular hospitalisation and re-hospitalisation, respectively (**Table-3**).

External validation

In our external validation cohort dataset, the incidence of cerebrovascular hospitalisation and re-hospitalisation were 4.99% and 1.87%, respectively. Applying the final models to our independent external cohort gave a C-statistic of 0.7098 (0.6875 to 0.7321) for cerebrovascular hospitalisation and 0.7184 (0.7041 to 0.7727) for cerebrovascular re-hospitalisation, and good calibration (bottom left of **Figure-1** for cerebrovascular hospitalisation and bottom right of **Figure-1** for cerebrovascular re-hospitalisation), with the calibration slope 0.9853 (0.9756 to 0.9966) and 0.9846 (0.8894 to 1.0796) for cerebrovascular hospitalisation and re-hospitalisation, respectively.

Clinical examples

Supplementary Chart-1 presents a real clinical example of the application of risk prediction model with graphical illustrations (risk score chart) for cerebrovascular hospitalisation and re-hospitalisation risk prediction scores to predict individual two-year risk of cerebrovascular hospitalisation and individual risk of re-hospitalisation within ninety days of a previous cerebrovascular hospitalisation.

DISCUSSION

Two new risk scores to quantify the individual absolute risk of cerebrovascular hospitalisation within two years and cerebrovascular re-hospitalisation after ninety days of prior cerebrovascular hospitalisation in a prospective cohort of type 2 diabetes patients in English primary care settings have been developed in this study. The two prediction models were validated externally in another independent prospectively English cohort. The two risk prediction scores revealed useful discrimination and excellent calibration, with C-statistics of bigger than 0.70 both in our derivation and external validation cohorts. The two risk prediction scores were derived from routine clinical measurements recorded and accessible in primary care settings, indicating that those can be applied in routine primary care (e.g. by embedding in practice software).

Kothari et al derived a prediction score to predict incident stroke within 10 years among 5,103 newly diagnosed type 2 diabetes patients in the UK Prospective Diabetes Study (UKPDS) [18]. Age, gender, atrial fibrillation, smoking, systolic blood pressure and lipid ratio were applied in the final model as predictors. However, the model performance (either discrimination or calibration) was not evaluated in the study. As the predictors like atrial fibrillation, duration of diabetes in the UKPDS algorithm were not available in our cohorts, we could not validate the UKPDS in our cohorts.

Yang et al derived a prediction model to predict incident stroke within 5 years among a Chinese diabetes population [19]. The splitting sample method was applied to the total sample (7920 type 2 diabetes patients) to generate a derivation sample (3,652 patients) and a validation sample (3,559 patients). The age, HbA1c, urinary albumin to creatinine ratio and history of coronary heart disease were included in the final model as predicators. The apparent C-statistics in the derivation sample was 0.78. And internal validated calibration suggested good. However, the splitting method was not suggested in the derivation of prediction models and the external validation was not implemented in this study. ***The source population in Yang's score was a Chinese population, which is different from our population (Caucasian population). The data from Yang's score were derived from a Diabetes registry (Hong Kong Diabetes Registry), which is different from our data source (primary care data). And the predictor "history of CHD" was not available in our cohorts. Therefore Yang's score could not be validated in our cohorts***

Previous risk prediction models have not addressed cerebrovascular disease as a group as a major reason and health cost for inpatient hospitalisation in type 2 diabetes patients. Being aware of the individual absolute risk of cerebrovascular hospitalisation in the following year, and the risk of a new episode (within ninety days) of a recurrent cerebrovascular event (re-hospitalisation) could help clinicians to process more intensive care to patients with a high risk profile and to decrease inpatient cost. Implementation approaches could be tested using a randomized controlled trial format including embedding alerts into practice software and increasing patient awareness of their risk.

There are several advantages in our two prediction models over those applied elsewhere. The two risk algorithms are on the basis of absolute risk derivation and validation in two prospective cohorts. Routine clinical measurements recorded in primary care settings were used to derive the two prediction models, which indicates that these measurements can be used straightforwardly in primary care and are modifiable for external validations in those developed countries that have primary care electronic health recorded dataset accessible for such objectives. The two scores can be readily imbedded into online tools for their application in primary care settings.

The approaches applied to develop and validate models are close to those models developed from the CPRD and QResearch studies [20, 21]. The predictors/parameters in the final scores are accurate and reliable clinical variables routinely recorded in general practices and routinely updated and reviewed for patients with type 2 diabetes, and are less varied than in other primary care electronic health record datasets. Moreover, the volume of missing values was relatively low, which would be less likely to lead to variation in potential external applications, although multiple imputation was applied.

We acknowledge that anti-diabetes treatments, diabetes duration, previous history of cerebrovascular diseases, other type 2 diabetes complications (e.g. renal failure), anti-hypertensive treatments, lifestyle relevant predictors (like smoking), and comorbidities were not taken into account due to limitations in our original data, but some prognostic factors were very common in people with diabetes (like antihypertensive treatments which is 81.2% in patients with type 2 diabetes in the United Kingdom[22]) which would be less discriminated in the model and we believe that the clinical measurements incorporated in the two prediction models could be proxies for inaccessible predictors. Data access limitations also barricaded extending the risk prediction model to all diabetes complications rather than those relevant to cerebrovascular hospitalisation. Due to the similarity between the development and validation cohort datasets, further more independent external validation (e.g. external data from more developed countries) are warranted.

To our knowledge, this is the 1st research to derive risk scores to quantify the two-year risk of cerebrovascular hospitalisation and re-hospitalisation within ninety days of a prior hospitalisation. For primary care practice these new two algorithms have two useful implications. First, these models can be use as screening tools to identify patients with high probability of cerebrovascular hospitalisation and re-hospitalisation. The two models are based on routine accessible clinical information recorded in primary care settings and evaluated by diabetes care teams. They can be imbedded into general practice computer systems or integrated into a mobile application for a handheld mobile device for ease of utilisation. Secondly, the risk scores could be applied to establish new thresholds of treatment in primary care practice through consensus development of guidance.

Conflict of interest

The authors declare that there is no conflict of interests with the publication of this paper.

Acknowledgments

This work was supported by the National Institute for Health Research under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0808-17303) and National Institute for Health Research for Patient Benefit Programme (Ref PB-PG-0610-22311). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

We would like to thank Simon Cohn, Chris Bunn, Toby Prevost, Charlotte Paddison, Sarah Donald, Candice Ward, Peers for Progress, West Anglia CLRN, Cambridgeshire & Peterborough PCT, Primary Care Research Network – East of England, MRC Epidemiology Unit, Eastern Diabetes Research Network, participating general practices, Jackie Williams, Caroline Taylor, Kym Mercer, Kevin Baker, Ben Bowers, Kalsoom Akhter (CUH Wolfson Diabetes & Endocrinology Clinic), James Brimicombe (Cambridge University), Kim Birch of Trumpington St General Practice, CUH Wolfson Diabetes & Endocrinology Clinic Educators, The RAPSID Patient Committee (Phillip Jones, Liz Carvlin, Roger Smith) and the peers and peer support participants.

References

1. Emerging Risk Factors Collaboration, Sarwar, N, Gao, P, Seshasai, SR, Gobin, R, Kaptoge, S et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet, 2010;9733:2215-2222.

2. Fox, CS, Coady, S, Sorlie, PD, D'Agostino RB, S, Pencina, MJ, Vasan, RS et al. Increasing cardiovascular disease burden due to diabetes mellitus: the Framingham Heart Study. Circulation, 2007;12:1544-1550.

3. Authors/Task Force Members, Ryden, L, Grant, PJ, Anker, SD, Berne, C, Cosentino, F et al. ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD: the Task Force on diabetes, pre-diabetes, and cardiovascular diseases of the European Society of Cardiology (ESC) and developed in collaboration with the European Association for the Study of Diabetes (EASD). Eur. Heart J., 2013;39:3035-3087.

4. Sampson, MJ, Dozio, N, Ferguson, B, Dhatariya, K. Total and excess bed occupancy by age, specialty and insulin use for nearly one million diabetes patients discharged from all English Acute Hospitals. Diabetes Res. Clin. Pract., 2007;1:92-98.

5. Hatzitolios, AI, Didangelos, TP, Zantidis, AT, Tziomalos, K, Giannakoulas, GA, Karamitsos, DT. Diabetes mellitus and cerebrovascular disease: which are the actual data?. J. Diabetes Complications., 2009;4:283-296.

6. Khalid, JM, Raluy-Callado, M, Curtis, BH, Boye, KS, Maguire, A, Reaney, M. Rates and risk of hospitalisation among patients with type 2 diabetes: retrospective cohort study using the UK General Practice Research Database linked to English Hospital Episode Statistics. Int. J. Clin. Pract., 2014;1:40-48.

7. van Dieren, S, Beulens, JW, Kengne, AP, Peelen, LM, Rutten, GE, Woodward, M et al. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. Heart, 2012;5:360-369.

8. Simmons, D, Prevost, AT, Bunn, C, Holman, D, Parker, RA, Cohn, S et al. Impact of community based peer support in type 2 diabetes: a cluster randomised controlled trial of individual and/or group approaches. PLoS One, 2015;3:e0120277.

9. Simmons, D, Cohn, S, Bunn, C, Birch, K, Donald, S, Paddison, C et al. Testing a peer support intervention for people with type 2 diabetes: a pilot for a randomised controlled trial. BMC Fam. Pract., 2013:5-2296-14-5.

10. Xie, B, Ye, XL, Sun, ZL, Jia, M, Jin, H, Ju, CP et al. Peer support for patients with type 2 diabetes in rural communities of China: protocol for a cluster randomized controlled trial. BMC Public Health, 2014:747-2458-14-747.

11. Smith, SM, Paul, G, Kelly, A, Whitford, DL, O'Shea, E, O'Dowd, T. Peer support for patients with type 2 diabetes: cluster randomised controlled trial. BMJ, 2011:d715.

12. Simmons, D, Yu, D, Wenzel, H. Changes in hospital admissions and inpatient tariff associated with a Diabetes Integrated Care Initiative: preliminary findings. J. Diabetes, 2014;1:81-89.

13. Yu, D, Simmons, D. Association between pulse pressure and risk of hospital admissions for cardiovascular events among people with Type 2 diabetes: a population-based case-control study. Diabet. Med., 2015;9:1201-1206.

14. Yu, D, Simmons, D. Association between blood pressure and risk of cardiovascular hospital admissions among people with type 2 diabetes. Heart, 2014;18:1444-1449.

15. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165:710-8.

16. Yu, D, Simmons, D. Association between pulse pressure and risk of hospital admissions for cardiovascular events among people with Type 2 diabetes: a population-based case-control study. Diabet. Med., 2015;9:1201-1206.

17. Moons, KG, Altman, DG, Reitsma, JB, Ioannidis, JP, Macaskill, P, Steyerberg, EW et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann. Intern. Med., 2015;1:W1-73.

18. Kothari, V, Stevens, RJ, Adler, AI, Stratton, IM, Manley, SE, Neil, HA et al. UKPDS 60: risk of stroke in type 2 diabetes estimated by the UK Prospective Diabetes Study risk engine. Stroke, 2002;7:1776-1781.

19. Yang, X, So, WY, Kong, AP, Ho, CS, Lam, CW, Stevens, RJ et al. Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. Diabetes Care, 2007;1:65-70.

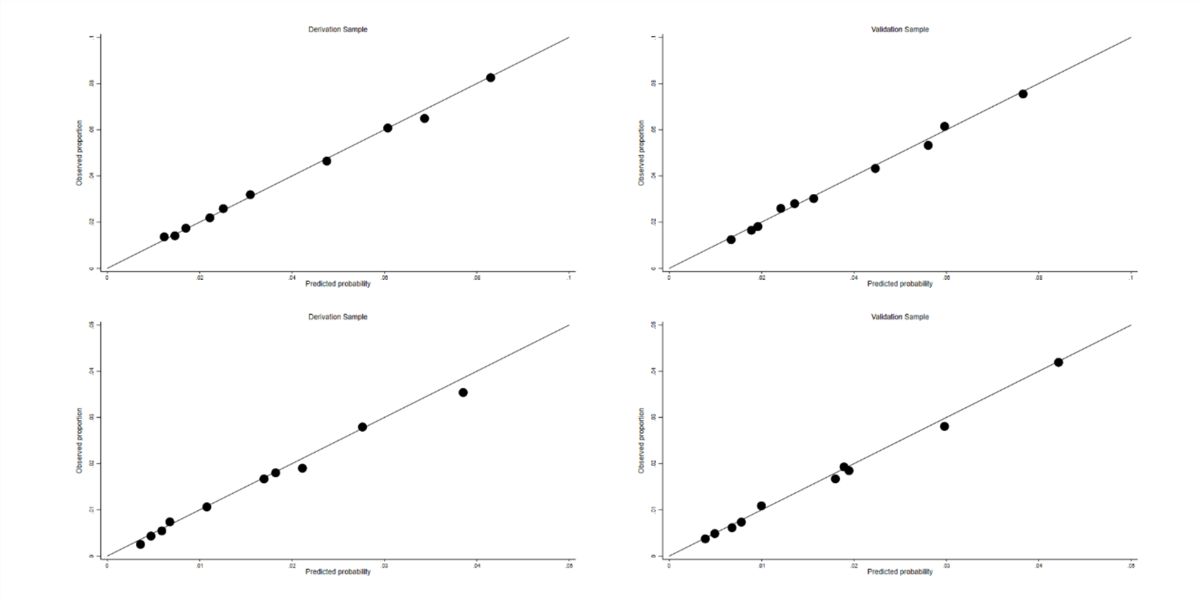
20. Nwaru, BI, Simpson, CR, Sheikh, A, Kotz, D. External validation of a COPD prediction model using population-based primary care data: a nested case-control study. Sci. Rep., 2017:44702.

21. Hippisley-Cox, J, Coupland, C, Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ, 2017:j2099.

22. Stone, MA, Charpentier, G, Doggen, K, Kuss, O, Lindblad, U, Kellner, C et al. Quality of care of people with type 2 diabetes in eight European countries: findings from the Guideline Adherence to Enhance Care (GUIDANCE) study. Diabetes Care, 2013;9:2628-2638.

FIGURE LEGENDS

Figure-1. Assessing calibration in the derivation cohort (left) and the validation cohort (right) for cerebrovascular hospitalisation (above panel) and cerebrovascular re-hospitalisation (below panel)



TABLES

Table-1. Characteristics of study participants in development cohort and external validation cohort.

	Development cohort	Validation cohort
Number of participants	4,704	1,121
Cerebrovascular hospitalisation, n (%)	244 (3.76)	56 (4.99)
Cerebrovascular rehospitalisation, n (%)	95 (1.46)	21 (1.87)
Age a baseline, years	65.0±16.3	65.5±11.4
Female gender, n (%)	1,919 (40.8)	444 (39.6)
Systolic blood pressure, mmHg	134.5±16.0	139.7±20.2
Diastolic blood pressure, mmHg	76.3±10.0	75.5±11.5
Total cholesterol, mmol/L	4.3±1.2	4.2±1.7
High density lipoprotein cholesterol, mmol/L	1.3±0.6	1.1±1.2
Low density lipoprotein cholesterol, mmol/L	2.5±1.4	1.4±3.0
Body mass index, kg/m^2	30.8±6.9	32.2±6.0
Glycated haemoglobin (HbA1c), mmol/mol / %	61.5±17.2 / 7.8±3.7	56.2±15.1 / 7.3±3.5
Taking lipid Lowering treatment, n (%)	3,342 (71.4)	731 (65.2)

Table-2. Multivariable model estimation for cerebrovascular hospitalisation and re-hospitalisation risk among type 2 diabetes patients in development cohort

Predictors/Parameters	**Coefficient**	**95% Confidence Interval**
Cerebrovascular Hospitalisation		
Male gender	0.3313	(0.2909 to 0.3716)
Glycated haemoglobin (HbA1c) ≥ 57 mmol/mol (7.4%)	-0.1259	(-0.1638 to -0.0879)
(Body mass index/10)^3	0.0624	(0.0520 to 0.0728)
((Body mass index/10)^3)*ln(Body mass index/10)	-0.0371	(-0.0435 to -0.0307)
Systolic blood pressure/100	1.6098	(0.4821 to 2.7375)
(Systolic blood pressure/100)^2	-0.2216	(-0.6220 to 0.1788)
(Diastolic blood pressure/100)^-2	-0.0239	(-0.0483 to 0.0005)
Diastolic blood pressure/100	-2.1136	(-2.3820 to -1.8452)
(Total cholesterol/10)^-2	-0.0056	(-0.0079 to -0.0033)
(Total cholesterol/10)^2	0.8866	(0.6862 to 1.0870)
(High density lipoprotein cholesterol)^3	0.0851	(0.0563 to 0.1139)
((High density lipoprotein cholesterol)^3)*ln(High density lipoprotein cholesterol)	-0.0892	(-0.1192 to -0.0593)
Low density lipoprotein cholesterol/10	-0.6356	(-0.9387 to -0.3325)
(Low density lipoprotein cholesterol/10)^3	0.5521	(-0.2076 to 1.3117)
Baseline age>=70 years	1.0647	(1.0213 to 1.1080)
Constant	-4.7571	(-5.5717 to -3.9426)
Cerebrovascular Re-hospitalisation		
Male gender	0.1359	(0.0741 to 0.1978)
Glycated haemoglobin (HbA1c) ≥ 57 mmol/mol (7.4%)	-0.2318	(-0.2914 to -0.1722)
(Body mass index/10)^3	0.0618	(0.0445 to 0.0792)
((Body mass index/10)^3)*ln(Body mass index/10)	-0.0383	(-0.0491 to -0.0274)
Systolic blood pressure/100	-2.4341	(-3.7885 to -1.0798)
(Systolic blood pressure/100)^2	1.2371	(0.7573 to 1.7169)
(Diastolic blood pressure/100)^-2	0.6846	(0.4897 to 0.8794)

((Diastolic blood pressure/100)^-2)*ln(Diastolic blood pressure/100)	0.3780	(0.2058 to 0.5501)
(Total cholesterol/10)^3	-1.4790	(-2.3056 to -0.6524)
((Total cholesterol/10)^3)*ln(Total cholesterol/10)	-11.2187	(-13.7345 to -8.7029)
(High density lipoprotein cholesterol)^3	0.1949	(0.1535 to 0.2362)
((High density lipoprotein cholesterol)^3)*ln(High density lipoprotein cholesterol)	-0.1992	(-0.2412 to -0.1572)
Low density lipoprotein cholesterol/10	0.0291	(-0.4999 to 0.5582)
(Low density lipoprotein cholesterol/10)^3	-1.6879	(-3.0975 to -0.2784)
Baseline age>=70 years	1.1117	(1.0424 to 1.1811)
Constant	-6.2027	(-7.2062 to -5.1991)

Table-3. Model performance statistics (with 95% confidence interval)

	Derivation				
Measure	Apparent performance	Test performance	Average optimism	Optimism corrected performance	External validation
	Cerebrovascular Hospitalisation				
C-statistic	1.0000 (0.9967 to 1.0034)	0.9961 (0.9884 to 1.0038)	0.0039	0.9961 (0.9928 to 0.9995)	0.9853 (0.9756 to 0.9966)
Calibration slope	0.7546 (0.7473 to 0.7619)	0.7509 (0.7454 to 0.7564)	0.0037	0.7509 (0.7436 to 0.7582)	0.7098 (0.6875 to 0.7321)
	Cerebrovascular Re-hospitalisation				
C-statistic	1.0000 (0.9557 to 1.0443)	0.9904 (0.9187 to 1.0621)	0.0096	0.9904 (0.9091 to 1.0525)	0.9846 (0.8894 to 1.0796)
Calibration slope	0.7476 (0.7403 to 0.7549)	0.7391 (0.7246 to 0.7536)	0.0085	0.7391 (0.7161 to 0.7451)	0.7184 (0.7041 to 0.7327)